



(12) 发明专利

(10) 授权公告号 CN 113223536 B

(45) 授权公告日 2024.04.19

(21) 申请号 202010062402.X

CN 110010133 A, 2019.07.12

(22) 申请日 2020.01.19

CN 110309880 A, 2019.10.08

(65) 同一申请的已公布的文献号

US 2017358306 A1, 2017.12.14

申请公布号 CN 113223536 A

US 2019117087 A1, 2019.04.25

(43) 申请公布日 2021.08.06

Cheng-I Lai et al..《ASSERT:Anti-Spoofing with Squeeze-Excitation and Residual neTworks》.《arXiv:1904.01120v1》.2019,第1-5页.

(73) 专利权人 TCL科技集团股份有限公司

地址 516006 广东省惠州市仲恺高新区惠风三路17号TCL科技大厦

Taejun Kim et al..《Comparison and Analysis of SampleCNN Architectures for Audio Classification》.《IEEE Journal of Selected Topics in Signal Processing》.2019,第13卷(第2期),第285-297页.

(72) 发明人 唐延欢

(74) 专利代理机构 深圳中一联合知识产权代理有限公司 44414

专利代理师 李木燕

Lei Fan et al..《Semantic Segmentation With Global Encoding and Dilated Decoder in Street Scenes》.《IEEE Access》.2018,第6卷第50333-50343页.

(51) Int.Cl.

G10L 17/00 (2013.01)

G10L 17/02 (2013.01)

G10L 17/18 (2013.01)

审查员 李海龙

(56) 对比文件

CN 107492382 A, 2017.12.19

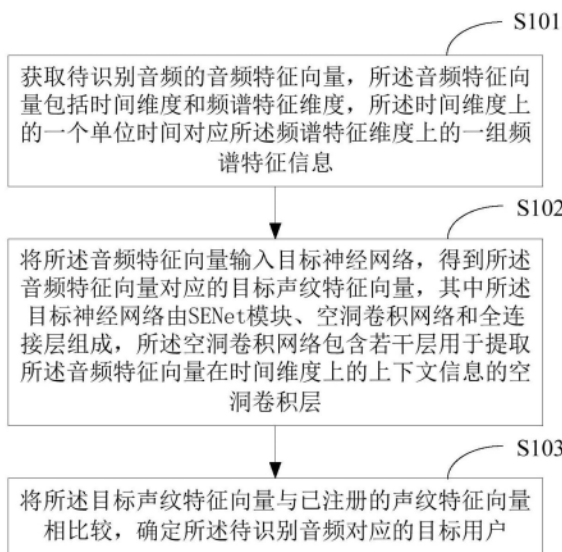
权利要求书3页 说明书13页 附图3页

(54) 发明名称

声纹识别方法、装置及终端设备

(57) 摘要

本申请适用于语音处理技术领域,提供了声纹识别方法、装置及终端设备,包括:获取待识别音频的音频特征向量;将所述音频特征向量输入目标神经网络,得到所述音频特征向量对应的目标声纹特征向量,其中所述目标神经网络由SENet模块、空洞卷积网络 and 全连接层组成,所述空洞卷积网络包含若干层用于提取所述音频特征向量在时间维度上的上下文信息的空洞卷积层;将所述目标声纹特征向量与已注册的声纹特征向量相比较,确定所述待识别音频对应的目标用户。本申请实施例能够在提高声纹识别的准确率的同时保证识别效率。



1. 一种声纹识别方法,其特征在于,包括:

获取待识别音频的音频特征向量,所述音频特征向量包括时间维度和频谱特征维度,所述时间维度上的一个单位时间对应所述频谱特征维度上的一组频谱特征信息;

将所述音频特征向量输入目标神经网络,得到所述音频特征向量对应的目标声纹特征向量,其中所述目标神经网络由SENet模块、空洞卷积网络 and 全连接层组成,所述空洞卷积网络包含若干层用于提取所述音频特征向量在时间维度上的上下文信息的空洞卷积层;

将所述目标声纹特征向量与已注册的声纹特征向量相比较,确定所述待识别音频对应的目标用户;

所述目标神经网络具体由第一卷积层、SENet模块、第一重构层、第一全连接层、第二重构层、空洞卷积网络、第三重构层、平均池化层和第二全连接层组成,所述将所述音频特征向量输入目标神经网络,得到所述音频特征向量对应的目标声纹特征向量,包括:

将所述音频特征向量输入目标神经网络,通过所述第一卷积层,得到第一特征向量,其中所述第一特征向量包含时间维度、频谱特征维度和通道维度;

所述第一特征向量通过所述SENet模块,对每个通道的信息进行加权,得到第二特征向量;

所述第二特征向量依次通过所述第一重构层、所述第一全连接层及所述第二重构层,得到第三特征向量;

所述第三特征向量通过所述空洞卷积网络,依次经过若干层空洞卷积层,提取所述第三特征向量在时间维度的上下文信息,得到第四特征向量,其中每层空洞卷积层包含一个尺寸为 $n \times 1$ 的卷积核, n 为大于1的正整数,“ \times ”为乘号;

所述第四特征向量依次通过第三重构层、平均池化层和第二全连接层,得到目标大小的目标声纹特征向量。

2. 如权利要求1所述的声纹识别方法,其特征在于,所述获取待识别音频的音频特征向量,包括:

获取待识别音频,滤除所述待识别音频的静音,得到有效音频段;并按照目标时长截取所述有效音频段,得到目标音频;

提取所述目标音频的梅尔倒谱系数MFCC特征,得到音频特征向量。

3. 如权利要求1所述的声纹识别方法,其特征在于,在所述获取待识别音频的音频特征向量之前,还包括:

获取样本数据,其中所述样本数据来自不同用户的音频数据;

将所述样本数据输入所述目标神经网络进行训练,直至类内音频相似度和类间音频相似度满足预设条件,得到训练后的目标神经网络;其中,所述类内音频相似度为属于同一用户的声纹特征向量之间的相似度,所述类间音频相似度为属于不同用户的声纹特征向量之间的相似度。

4. 如权利要求3所述的声纹识别方法,其特征在于,所述将所述样本数据输入所述目标神经网络进行训练,直至类内音频相似度和类间音频相似度满足预设条件,得到训练后的目标神经网络,包括:

依次将预设样本数据输入所述目标神经网络进行训练,直至目标函数的值满足预设条件,得到训练后的目标神经网络,所述目标神经网络的目标函数为:

$$Sc = \sum_i^{NM} (\sum_j^{M-1} \text{sim}\{(v_i, v_j) | i \neq j, v_i \in P, v_j \in P\} - \max(\text{sim}\{(v_i, v_k) | v_i \in P, v_k \notin P\}))$$

其中, Sc 为目标函数的值, 表示类内音频相似度与类间音频相似度之间的差值; N 为当前批次输入的样本数据对应的用户的数目, M 为每个用户对应的样本数据数量, v_i 表示当前批次的任意一个样本数据经过目标神经网络模型得到的声纹特征向量, P 表示 v_i 对应的用户, v_j 为与 v_i 同属于一个用户的声纹特征向量, v_k 为与 v_i 不属于同一个用户的声纹特征向量, $\text{sim}\{(v_i, v_j) | i \neq j, v_i \in P, v_j \in P\}$ 表示 v_i 与同属于 P 的声纹特征向量 v_j 的余弦相似度, $\text{sim}\{(v_i, v_k) | v_i \in P, v_k \notin P\}$ 表示 v_i 与其他用户的声纹特征向量 v_k 的余弦相似度。

5. 如权利要求3所述的声纹识别方法, 其特征在于, 所述目标神经网络在训练时的学习率根据预设的目标学习率和当前训练步数动态调整。

6. 如权利要求1至5任意一项所述的声纹识别方法, 其特征在于, 在所述将所述目标声纹特征向量与已注册的声纹特征向量相比较之后, 包括:

若未找到与所述目标声纹特征向量相匹配的已注册的声纹特征向量, 则指示当前用户将所述目标声纹特征向量进行注册。

7. 一种声纹识别装置, 其特征在于, 包括:

音频特征向量获取单元, 用于获取待识别音频的音频特征向量, 所述音频特征向量包括时间维度和频谱特征维度, 所述时间维度上的一个单位时间对应所述频谱特征维度上的一组频谱特征信息;

目标神经网络单元, 用于将所述音频特征向量输入目标神经网络, 得到所述音频特征向量对应的目标声纹特征向量, 其中所述目标神经网络由SENet模块、空洞卷积网络和全连接层组成, 所述空洞卷积网络包含若干层用于提取所述音频特征向量在时间维度上的上下文信息的空洞卷积层;

确定单元, 用于将所述目标声纹特征向量与已注册的声纹特征向量相比较, 确定所述待识别音频对应的目标用户;

所述目标神经网络具体由第一卷积层、SENet模块、第一重构层、第一全连接层、第二重构层、空洞卷积网络、第三重构层、平均池化层和第二全连接层组成, 所述目标神经网络单元具体用于:

将所述音频特征向量输入目标神经网络, 通过所述第一卷积层, 得到第一特征向量, 其中所述第一特征向量包含时间维度、频谱特征维度和通道维度;

所述第一特征向量通过所述SENet模块, 对每个通道的信息进行加权, 得到第二特征向量;

所述第二特征向量依次通过所述第一重构层、所述第一全连接层及所述第二重构层, 得到第三特征向量;

所述第三特征向量通过所述空洞卷积网络, 依次经过若干层空洞卷积层, 提取所述第三特征向量在时间维度的上下文信息, 得到第四特征向量, 其中每层空洞卷积层包含一个尺寸为 $n \times 1$ 的卷积核, n 为大于1的正整数, “*” 为乘号;

所述第四特征向量依次通过第三重构层、平均池化层和第二全连接层, 得到目标大小的目标声纹特征向量。

8. 一种终端设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现如权利要求1至6任一项所述方法的步骤。

9. 一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至6任一项所述方法的步骤。

声纹识别方法、装置及终端设备

技术领域

[0001] 本申请属于语音处理技术领域,尤其涉及一种声纹识别方法、装置及终端设备。

背景技术

[0002] 声纹识别 (Voiceprint Recognition, VPR) 也称为说话人识别 (Speaker Recognition), 作为生物识别技术的一种, 长久以来受到学术界、工业界的广泛关注。传统的声纹识别技术以 i-vectors 为经典, 但其准确性较差。为此, 谷歌提出 GE2E (Generalized end-to-end) 网络结构, 在声纹识别中, 其比 i-vectors 拥有更高的识别准确率。然而, GE2E 复杂的神经网络结构, 使其模型占用空间过大、识别速度缓慢, 不利于实际生产环境的应用。

发明内容

[0003] 有鉴于此, 本申请实施例提供了声纹识别方法、装置及终端设备, 以解决现有技术中如何在提高声纹识别的准确率的同时保证识别效率的问题。

[0004] 本申请实施例的第一方面提供了一种声纹识别方法, 包括:

[0005] 获取待识别音频的音频特征向量, 所述音频特征向量包括时间维度和频谱特征维度, 所述时间维度上的一个单位时间对应所述频谱特征维度上的一组频谱特征信息;

[0006] 将所述音频特征向量输入目标神经网络, 得到所述音频特征向量对应的目标声纹特征向量, 其中所述目标神经网络由 SENet 模块、空洞卷积网络和全连接层组成, 所述空洞卷积网络包含若干层用于提取所述音频特征向量在时间维度上的上下文信息的空洞卷积层;

[0007] 将所述目标声纹特征向量与已注册的声纹特征向量相比较, 确定所述待识别音频对应的目标用户。

[0008] 本申请实施例的第二方面提供了一种声纹识别装置, 包括:

[0009] 音频特征向量获取单元, 用于获取待识别音频的音频特征向量, 所述音频特征向量包括时间维度和频谱特征维度, 所述时间维度上的一个单位时间对应所述频谱特征维度上的一组频谱特征信息;

[0010] 目标神经网络单元, 用于将所述音频特征向量输入目标神经网络, 得到所述音频特征向量对应的目标声纹特征向量, 其中所述目标神经网络由 SENet 模块、空洞卷积网络和全连接层组成, 所述空洞卷积网络包含若干层用于提取所述音频特征向量在时间维度上的上下文信息的空洞卷积层;

[0011] 确定单元, 用于将所述目标声纹特征向量与已注册的声纹特征向量相比较, 确定所述待识别音频对应的目标用户。

[0012] 本申请实施例的第三方面提供了一种终端设备, 包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序, 所述处理器执行所述计算机程序时实现如所述声纹识别方法的步骤。

[0013] 本申请实施例的第四方面提供了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现如所述声纹识别方法的步骤。

[0014] 本申请实施例的第五方面提供了一种计算机程序产品,当计算机程序产品在终端设备上运行时,使得终端设备执行上述声纹识别方法的。

[0015] 本申请实施例与现有技术相比存在的有益效果是:本申请实施例中,将待识别音频的音频特征向量通过由SENet模块、空洞卷积网络和全连接层组成的目标神经网络进行特征提取,得到目标声纹特征向量,并与已注册的声纹特征向量相比较,确定待识别音频对应的目标用户,由于SENet模块能够加强通道间的特征信息提取,并且空洞卷积网络中能够提取待识别音频时间维度上的上下文信息,因此能够使得最终提取的目标声纹特征向量的包含的特征信息更加准确全面,从而使得声纹识别更加准确;同时,由于该目标神经网络相对于GE2E网络的结构简单,因此能够降低提取声纹特征信息时的复杂度,提高声纹特征信息提取的效率,从而提高声纹识别的效率。

附图说明

[0016] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0017] 图1是本申请实施例提供的第一种声纹识别方法的实现流程示意图;

[0018] 图2是本申请实施例提供的一种目标神经网络的结构示意图;

[0019] 图3是本申请实施例提供的第二种声纹识别方法的实现流程示意图;

[0020] 图4是本申请实施例提供的声纹识别装置的示意图;

[0021] 图5是本申请实施例提供的终端设备的示意图。

具体实施方式

[0022] 以下描述中,为了说明而不是为了限定,提出了诸如特定系统结构、技术之类的具体细节,以便透彻理解本申请实施例。然而,本领域的技术人员应当清楚,在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中,省略对众所周知的系统、装置、电路以及方法的详细说明,以免不必要的细节妨碍本申请的描述。

[0023] 为了说明本申请所述的技术方案,下面通过具体实施例来进行说明。

[0024] 应当理解,当在本说明书和所附权利要求书中使用时,术语“包括”指示所描述特征、整体、步骤、操作、元素和/或组件的存在,但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

[0025] 还应当理解,在此本申请说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本申请。如在本申请说明书和所附权利要求书中所使用的那样,除非上下文清楚地指明其它情况,否则单数形式的“一”、“一个”及“该”意在包括复数形式。

[0026] 还应当进一步理解,在本申请说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合,并且包括这些组合。

[0027] 如在本说明书和所附权利要求书中所使用的那样,术语“如果”可以依据上下文被解释为“当...时”或“一旦”或“响应于确定”或“响应于检测到”。类似地,短语“如果确定”或“如果检测到[所描述条件或事件]”可以依据上下文被解释为意指“一旦确定”或“响应于确定”或“一旦检测到[所描述条件或事件]”或“响应于检测到[所描述条件或事件]”。

[0028] 另外,在本申请的描述中,术语“第一”、“第二”、“第三”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0029] 实施例一:

[0030] 图1示出了本申请实施例提供的第一种声纹识别方法的流程示意图,详述如下:

[0031] 在S101中,获取待识别音频的音频特征向量,所述音频特征向量包括时间维度和频谱特征维度,所述时间维度上的一个单位时间对应所述频谱特征维度上的一组频谱特征信息。

[0032] 本申请实施例中的声纹识别方法具体为文本无关的声纹识别方法,即用户无需按照规定的内容发音,本申请实施例的待识别音频为用户发出的任意说话内容的音频。通过声音采集设备或者存储待识别音频的存储单元中获取待识别音频,并通过时域、频率域的变换、分析,提取所述待识别音频的音频特征向量,该音频特征向量可用numpy文件格式进行存储。其中,音频特征向量包括时间维度和频谱特征维度,音频特征向量可用 $a*b$ 表示,其中“*”为乘号, a 为音频特征向量在时间维度上的长度, b 为音频特征向量在频谱特征维度上的长度,时间维度上的一个单位时间对应频谱特征维度上的一组频谱特征信息,即音频特征向量包含了 a 个单位时间的音频特征信息,每个单位时间的音频特征信息可以用一组长度为 b 的频谱特征信息表示。

[0033] 具体地,所述步骤S101具体包括:

[0034] 获取待识别音频,滤除所述待识别音频的静音,得到有效音频段;并按照目标时长截取所述有效音频段,得到目标音频;

[0035] 提取所述目标音频的梅尔倒谱系数MFCC特征,得到音频特征向量。

[0036] 获取待识别音频,并滤除该待识别音频的静音,得到该待识别音频的有效音频段,并按照目标时长(例如3秒)截取该待识别音频的有效音频段,得到目标音频。可选地,若该待识别音频的有效音频段的时长小于有效时长阈值(例如1秒),则丢弃该待识别音频并重新获取待识别音频。可选地,若该待识别音频的有效音频段的时长大于或者等于有效时长阈值且小于或者目标时长时,则不对所述待识别音频进行截取,直接以该有效音频段作为目标音频,因此本申请实施例的目标音频的时长大于或者等于有效时长阈值,小于或者等于目标时长。具体地,所述目标时长或者有效时长阈值根据目标神经网络的识别精度确定。可选地,本申请实施例的待识别音频具体为短音频(例如时长在1-3秒的音频),本申请实施例的目标神经网络只需根据较少的短音频的音频特征信息便可准确地进行声纹识别,即由于本申请实施例的目标神经网络识别准确度高,所需的输入信息量较少,因此获取的待识别音频的时长可以缩短,并使得目标神经网络处理的数据量也减少,从而能够提高声纹识别的效率。

[0037] 将所述目标音频按照预设的采样率、分帧帧长、第一步长等参数进行梅尔倒谱系数(Mel-scale Frequency Cepstral Coefficients, MFCC)特征提取,得到包含时间维度和频谱特征维度的音频特征向量,该音频特征向量的尺寸为规定尺寸,即在时间维度上的长

度和频谱特征维度上的长度都为根据MFCC特征提取时的参数设置得到的目标长度。作为示例而非限定,预设的采样率为1.6k,分帧帧长为25毫秒、第一步长为32毫秒,音频特征向量的尺寸为96*64。可选地,若目标音频的时长小于3秒,则通过MFCC特征提取后得到的第一音频特征向量的尺寸小于规定尺寸,此时将第一音频特征向量中不足规定尺寸的部分用“0”进行数据填充,补齐得到规定尺寸的第二音频特征向量作为最终输入目标神经网络的音频特征向量。例如,设目标音频的时长为2秒,该目标音频通过MFCC特征提取后得到的第一音频特征向量的尺寸为63*64,该尺寸小于规定尺寸“96*64”,因此将第一音频特征向量中与规定尺寸相差的“33*64”尺寸的位置用“0”进行数据填充,补齐得到尺寸为96*64的第二音频特征向量作为最终的音频特征向量。

[0038] 在S102中,将所述音频特征向量输入目标神经网络,得到所述音频特征向量对应的目标声纹特征向量,其中所述目标神经网络由SENet模块、空洞卷积网络和全连接层组成,所述空洞卷积网络包含若干层用于提取所述音频特征向量在时间维度上的上下文信息的空洞卷积层。

[0039] 将规定尺寸的音频特征向量通过单通道(即输入通道数为1)输入目标神经网络,对音频特征向量进行进一步特征提取处理,得到音频特征向量对应的目标声纹特征向量。目标神经网络由Squeeze-and-Excitation Networks(简称SENet)模块、空洞卷积网络和全连接层组成,SENet模块用于对目标神经网络中各通道间的相关性进行建模,SENet在训练时根据样本数据确定了每个通道的权重参数,使得之后通过训练好的SENet模块能够按照该权重准确地提取各通道的特征信息从而提升声纹识别的准确率,空洞卷积网络用于提取音频特征向量在时间维度上的上下文信息。本申请实施例中,该上下文信息具体为融合了音频特征向量中前后多个不同单位时间对应的频谱特征信息的特征信息。具体地,该空洞卷积网络包含若干层空洞卷积层,每层空洞卷积层包含一个尺寸为 $n \times 1$ 的卷积核,其中 n 为大于1的正整数,“*”为乘号。通过该 $n \times 1$ 的卷积核针对整合分析音频特征向量在 n 个时间维度上的上下文信息,即加强音频特征提取过程中的上下文联系,从而提高声纹识别的准确率。并且,这里采用空洞卷积网络可以在不损失数据精确性的情况下,使得卷积核视野更加广阔。

[0040] 具体地,所述目标神经网络具体由第一卷积层、SENet模块、第一重构层、第一全连接层、第二重构层、空洞卷积网络、第三重构层、平均池化层和第二全连接层组成,所述步骤S102,具体包括:

[0041] S10201:将所述音频特征向量输入目标神经网络,通过所述第一卷积层,得到第一特征向量,其中所述第一特征向量包含时间维度、频谱特征维度和通道维度;

[0042] S10202:所述第一特征向量通过所述SENet模块,对每个通道的信息进行加权,得到第二特征向量;

[0043] S10203:所述第二特征向量依次通过所述第一重构层、所述第一全连接层及所述第二重构层,得到第三特征向量;

[0044] S10204:所述第三特征向量通过所述空洞卷积网络,依次经过若干层空洞卷积层,提取所述第三特征向量在不同时间维度的上下文信息,得到第四特征向量,其中每层空洞卷积层包含一个尺寸为 $n \times 1$ 的卷积核, n 为大于1的正整数,“*”为乘号;

[0045] S10205:所述第四特征向量依次通过第三重构层、平均池化层和第二全连接层,得

到目标大小的目标声纹特征向量。

[0046] 如图2所示,本申请实施例的目标神经网络由第一卷积层Conv1-ReLU、SENet模块、第一重构层Reshape1、第一全连接层Fc1、第二重构层Reshape2、空洞卷积网络Dilated-Conv-Net、第三重构层Reshape3、平均池化层Avg-pool和第二全连接层Fc2组成。

[0047] 在S10201中,将音频特征向量经单通道输入目标神经网络(即输入通道数为1),通过第一卷积层Conv1-ReLU,得到第一特征向量。该第一卷积层包含第一数量的通道,每个通道的卷积核为 3×3 ,步长为第二步长,相应地,经过第一卷积层输出的第一特征向量除了时间维度、频谱特征维度外,还包括通道维度,第一特征向量在通道维度上的长度等于第一数量。示例性地,所述音频特征向量的大小为 96×64 ,输入通道数为1,即可视为输入为 $96 \times 64 \times 1$ 的输入数据;第一卷积层有32个通道,每个通道的卷积核大小为 3×3 、步长为2,则经过第一卷积层后,得到尺寸为 $48 \times 32 \times 32$ 的第一特征向量,其中“48”为时间维度上的长度,第一个“32”为频谱特征维度上的长度,第二个“32”为通道维度上的长度。

[0048] 在S10202中,将第一卷积层输出的第一特征向量通过SENet模块,根据SENet的各通道权重参数对第一特征向量每个通道的信息进行加权,得到第二特征向量。第二特征向量的维度和尺寸与第一特征向量一致。

[0049] 在S10203中,将经过通道加权后得到的第二特征向量经过第一重构层Reshape1、第一全连接层Fc1和第二重构层Reshape2,得到通道维度上长度为1的第三特征向量。示例性地,第二特征向量的尺寸为 $48 \times 32 \times 32$,经过第一重构层Reshape1后得到尺寸为 $48 \times (32 \times 32) = 48 \times 1024$ 的特征向量;之后通过由全连接层和激活函数tanh组成的第一全连接层Fc1,映射得到尺寸为 48×256 的特征向量;接着通过第二重构层Reshape2扩充通道维度,得到尺寸为 $48 \times 256 \times 1$ 的第三特征向量。通过第一重构层、第一全连接层、第二重构层,将第二特征向量的数据进行样式调整得到单通道(通道维度上的长度为1)的第三特征向量,以适应空洞卷积网络对输入数据的格式要求。

[0050] S10204:所述第三特征向量通过所述空洞卷积网络,依次经过若干层空洞卷积层,提取所述第三特征向量在时间维度的上下文信息,得到第四特征向量;

[0051] 将第三特征向量经过空洞卷积网络中若干层包含 $n \times 1$ 的卷积核的空洞卷积层的处理,提取第三特征向量中时间维度上的上下文信息,即将每一个单位时间对应的频谱特征信息与相邻单位时间对应的频谱特征信息进行关联处理,得到包含时间维度上的上下文信息的第四特征向量,第四特征向量的尺寸大小与第三特征向量一致。具体地,该空洞卷积网络包含若干层空洞卷积层,每层空洞卷积层包含一个尺寸为 $n \times 1$ 的卷积核,其中n为大于1的正整数,“*”为乘号。通过该 $n \times 1$ 的卷积核针对整合分析音频特征向量在n个时间维度上的上下文信息,即加强音频特征提取过程中的上下文联系,从而提高声纹识别的准确率。示例性地,所述空洞卷积网络由Dilated-Conv1、Dilated-Conv2、Dilated-Conv3、Dilated-Conv4、Dilated-Conv5这五层空洞卷积层组成,这五层空洞卷积层均为单通道,步长均为1,卷积核大小依次为 5×1 、 9×1 、 15×1 、 24×1 、 24×1 ,相应的空洞卷积比例为1、2、3、1、1;设第三特征向量的尺寸为 $48 \times 256 \times 1$,则经过空洞卷积网络处理后的第四特征向量的尺寸也为 $48 \times 256 \times 1$ 。

[0052] 将第四特征向量依次通过第三重构层Reshape3进行数据样式调整,并通过平均池化层Avg-pool在时间维度上对不同单位时间的各频谱特征信息取平均值,得到时间维度上长度为1(即单时间维度,因此可省去时间维度上的表示)的特征向量;之后通过由全连接层

和tanh激活函数组成的第二全连接层,得到目标大小的目标声纹特征向量。示例性地,尺寸为48*256*1的第四特征向量通过第三重构层得到尺寸为48*256的特征向量,并通过平均池化层对48个不同单位时间的频谱特征信息相加求平均值,得到尺寸为256的特征向量(即频谱特征维度上的长度为256,时间维度和通道维度均被归一化的特征向量);之后通过第二全连接层,将大小为256的特征向量映射为目标大小为512的目标声纹特征向量,即该目标声纹特征向量包含512个单位的特征信息。

[0053] 在S103中,将所述目标声纹特征向量与已注册的声纹特征向量相比较,确定所述待识别音频对应的目标用户。

[0054] 本申请实施例中,预存了已注册的声纹特征向量及对应的用户的标识信息,该标识信息可以为用户的姓名、编号等信息。将步骤S102中得到的目标声纹特征向量和已注册的声纹特征向量相比较,找出与该目标声纹特征向量相似度最高的已注册的声纹特征向量,并确定该已注册的声纹向量对应的用户即为本次待识别音频对应的目标用户,输出该目标用户的标识信息。具体地,通过求目标声纹特征向量与预存的各个已注册的声纹特征向量的余弦相似度,找出与该目标声纹特征向量相似度最高的已注册的声纹特征向量。

[0055] 可选地,在步骤S101之前,包括:

[0056] 接收注册指令,获取待注册的用户标识信息及对应的待注册音频;

[0057] 通过目标神经网络获得该待注册音频的声纹特征向量;

[0058] 将该待注册音频的声纹特征向量和对应的用户标识信息存储至目标数据库,得到已注册的声纹特征向量及对应的用户的标识信息。

[0059] 优选地,在注册时,获取同一待注册的用户的多条待注册音频,将这多条待注册音频同时或者先后通过目标神经网络,相应获得同一待注册的用户多个声纹特征向量,并求这多个声纹特征向量的均值作为该待注册的用户最终的声纹特征向量进行注册,从而进一步提高注册数据的准确性,提高之后声纹识别的准确率。

[0060] 可选地,在所述步骤S103之后,还包括:

[0061] 若未找到与目标声纹特征向量相匹配的已注册的声纹特征向量,则指示当前用户将所述目标声纹特征向量进行注册。

[0062] 若没有找到和目标声纹特征向量相匹配的已注册的声纹特征向量,说明当前的待识别音频对应的用户信息尚未注册过,因此指示用户输入当前用户的标识信息,并将该用户的标识信息和目标声纹特征向量对应存储至目标数据库,完成目标声纹特征向量的注册。

[0063] 本申请实施例中,将待识别音频的音频特征向量通过由SENet模块、空洞卷积网络和全连接层组成的目标神经网络进行特征提取,得到目标声纹特征向量,并与已注册的声纹特征向量相比较,确定待识别音频对应的目标用户,由于SENet模块能够加强通道间的特征信息提取,并且空洞卷积网络中能够提取待识别音频时间维度上的上下文信息,因此能够使得最终提取的目标声纹特征向量的包含的特征信息更加准确全面,从而使得声纹识别更加准确;同时,由于该目标神经网络相对于GE2E网络的结构简单,因此能够降低提取声纹特征信息时的复杂度,提高声纹特征信息提取的效率,从而提高声纹识别的效率。

[0064] 应理解,上述实施例中各步骤的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定,而不对本申请实施例的实施过程构成任何限

定。

[0065] 实施例二:

[0066] 图3示出了本申请实施例提供的第二种声纹识别方法的流程示意图,详述如下:

[0067] 在S301中,获取样本数据,其中所述样本数据来自不同用户的音频数据。

[0068] 通过获取不同用户的音频数据并对音频数据进行预处理得到不同用户的音频特征向量作为样本数据。或者,通过读取以npy文件形式预存的不同用户的音频数据的音频特征向量,得到训练用的样本数据。具体地,该样本数据中,每个用户均存在两个或者两个以上的音频特征向量。

[0069] 在S302中,将所述样本数据输入所述目标神经网络进行训练,直至类内音频相似度和类间音频相似度满足预设条件,得到训练后的目标神经网络;其中,所述类内音频相似度为属于同一用户的不同音频数据对应的声纹特征向量之间的相似度,所述类间音频相似度为属于不同用户的不同音频数据对应的声纹特征向量之间的相似度。

[0070] 将样本数据输入目标神经网络进行训练,通过调整各网络层的学习参数,直至根据样本数据得到的声纹特征向量中,类内音频相似度和类间音频相似度满足预设条件,使得类内音频相似度尽量大,同时类间音频相似度尽量小。其中,类内音频相似度指的是属于同一用户的声纹特征向量之间的相似度,类间音频相似度为属于不同用户的声纹特征向量之间的相似度。具体地,声纹特征向量之间的相似度可以用余弦相似度表示。具体地,该预设条件可以为类内音频相似度大于第一预设阈值且类间音频相似度小于第二预设阈值,或者该预设条件可以为:类内音频相似度与类间音频相似度的差值大于预设差值,从而使得类内音频相似度尽量大,同时类间音频相似度尽量小。

[0071] 可选地,所述步骤S103包括:

[0072] 依次将预设样本数据输入所述目标神经网络进行训练,直至目标函数的值满足预设条件,得到训练后的目标神经网络,所述目标神经网络的目标函数为:

$$[0073] \quad Sc = \sum_i^{NM} \left(\sum_j^{M-1} \text{sim}\{(v_i, v_j) | i \neq j, v_i \in P, v_j \in P\} - \max(\text{sim}\{(v_i, v_k) | v_i \in P, v_k \notin P\}) \right)$$

[0074] 其中,Sc为目标函数的值,表示类内音频相似度与类间音频相似度之间的差值;N为当前批次输入的样本数据对应的用户的数目,M为每个用户对应的样本数据数量, v_i 表示当前批次的任意一个样本数据经过目标神经网络模型得到的声纹特征向量,P表示 v_i 对应的用户, v_j 为与 v_i 同属于一个用户的声纹特征向量, v_k 为与 v_i 不属于同一个用户的声纹特征向量, $\text{sim}\{(v_i, v_j) | i \neq j, v_i \in P, v_j \in P\}$ 表示 v_i 与同属于P的声纹特征向量 v_j 的余弦相似度, $\text{sim}\{(v_i, v_k) | v_i \in P, v_k \notin P\}$ 表示 v_i 与其他用户的声纹特征向量 v_k 的余弦相似度。

[0075] 每次从数据集中获取预设批次样本数目的样本数据输入目标神经网络进行训练,其中每批次的样本数据来自预设人数的用户各自对应的预设句数的音频。例如设预设批次样本数目为64,即每次输入64个样本数据作为一个批次对目标神经网络进行训练,这64个样本数据来源于16个用户,每个用户对应4句音频,即每个用户对应4个样本数据。

[0076] 目标神经网络的目标函数为

$$Sc = \sum_i^{NM} \left(\sum_j^{M-1} \text{sim}\{(v_i, v_j) | i \neq j, v_i \in P, v_j \in P\} - \max(\text{sim}\{(v_i, v_k) | v_i \in P, v_k \notin P\}) \right) \quad \text{其中,目标函数}$$

计算得到的值 S_c 表示类内音频相似度与类间音频相似度之间的差值; N 为当前批次输入的样本数据对应的用户的数目, M 为每个用户对应的样本数据数量, NM 表示 N 乘以 M ,等于预设批次样本数目。 v_i 表示当前批次的任意一个样本数据经过目标神经网络模型得到的声纹特征向量, P 表示 v_i 对应的用户, v_j 为与 v_i 同属于一个用户的声纹特征向量, v_k 为与 v_i 不属于同一个用户的声纹特征向量, $\text{sim}\{(v_i, v_j) \mid i \neq j, v_i \in P, v_j \in P\}$ 表示 v_i 与同属于 P 的声纹特征向量 v_j 的余弦相似度, $\text{sim}\{(v_i, v_k) \mid v_i \in P, v_k \notin P\}$ 表示 v_i 与其他用户的声纹特征向量 v_k 的余弦相似度。在训练时,对 S_c 取负值,采用梯度下降法训练直至 $(-S_c)$ 的值的下降梯度小于预设值且目标神经网络的准确率高于一准确率阈值,得到训练后的目标神经网络。即,本申请实施例的预设条件可以为“ $(-S_c)$ 的值的下降梯度小于预设值且目标神经网络的准确率高于一准确率阈值”,此时经过目标神经网络处理的音频中,类内音频相似度大且类间音频相似度小。优选地,当 $(-S_c)$ 取得最小值,即 S_c 取值最大,此时类内样本数据(即属于同一个用户的样本数据)的声纹特征向量之间的余弦相似度尽量大,同时类间样本数据(即分别属于不同用户的样本数据)的声纹特征向量之间的余弦相似度尽量小,对应的目标神经网络的识别准确度最高。

[0077] 可选地,所述目标神经网络在训练时的学习率根据预设的目标学习率和当前训练步数动态调整。

[0078] 具体地,目标神经网络在训练时的学习率根据预设的目标学习率和当前训练步数,通过warm up和学习率衰减相结合的方式,进行动态调整。具体地,训练时的学习率 lr 通过以下的学习率公式动态调整:

$$[0079] \quad lr = flr \times 10^{0.5} \times \min(\text{step} \times 10^{-1.5}, \text{step}^{-0.5})$$

[0080] 其中 flr 为预设的目标学习率, step 为当前训练步数。

[0081] 根据该学习率公式,学习率在训练初期时逐渐预热,增大至预设的目标学习率,加快训练收敛速度;在学习率到达目标学习率之后的训练后期阶段,学习率又逐渐衰减,使得目标神经网络能够更准确地收敛。通过该动态调整方式,能够提高目标神经网络的训练速度及准确度。

[0082] 可选地,所述声纹识别方法具体应用于远场录音场景,所述样本数据包括携带背景噪声的远场录音数据、预设数量的无噪声音频数据。

[0083] 本申请实施例的声纹识别方法具体应用于远场录音场景,例如智能电视的远场录音场景。在远场录音场景中的音频包含一定的背景噪声,相应地,训练目标神经网络的样本数据也包括含有背景噪声的远场录音数据。另外,由于远场录音数据可能过于嘈杂,导致目标神经网络比较难以收敛,因此本申请实施例中的样本数据除了含有背景噪声的远场录音数据,还包括预设数量的无噪声音频数据。将含有背景噪声的远场录音数据和预设数量的无噪声音频数据合并作为样本数据对目标神经网络进行训练,能够在保证训练的目标神经网络准确地贴合远场录音场景下的声纹识别的同时,提高目标神经网络的收敛速度。示例性地,本申请实施例中包括样本数据集包括来自5512个用户的共16074个远场录音数据(每个远场录音数据存储存储在 npv 文件中)以及来自2500个用户的255763个无噪声音频数据(每个无噪声音频数据存储存储在 npv 文件中)。

[0084] 在S303中,获取待识别音频的音频特征向量,所述音频特征向量包括时间维度和频谱特征维度,时间维度上的一个单位时间对应频谱特征维度上的一组频谱特征信息。

[0085] 在S304中,将所述音频特征向量输入目标神经网络,得到所述音频特征向量对应的目标声纹特征向量,其中所述目标神经网络由SENet模块、空洞卷积网络和全连接层组成,所述空洞卷积网络包含若干层用于提取所述音频特征向量在不同时间维度上的上下文信息的空洞卷积层。

[0086] 在S305中,将所述目标声纹特征向量与已注册的声纹特征向量相比较,确定所述待识别音频对应的目标用户。

[0087] 本申请实施例的S303~S305分别与上一实施例的S101~S103相同,具体请参阅上一实施例中S101~S103的相关描述,此处不赘述。

[0088] 本申请实施例中,通过对目标神经网络进行训练,直至类内音频相似度和类间音频相似度满足预设条件,使得最终训练得到的目标神经网络能够使类内样本数据(即属于同一个用户的样本数据)的声纹特征向量之间的余弦相似度尽量大,同时类间样本数据(即分别属于不同用户的样本数据)的声纹特征向量之间的余弦相似度尽量小,提高目标神经网络的识别准确度,从而提高声纹识别方法的准确度。

[0089] 作为示例而非限定,以下提供了本申请实施例的声纹识别方法的测试验证过程及结果:

[0090] (一) 准确率测试

[0091] A1: 获取样本数据集以外的来自6个用户、每人2句的语音数据,并进行预处理、MFCC特征提取,得到12个音频特征向量,每个音频特征向量均携带对应的用户标识信息;

[0092] A2: 将步骤A1中的12个音频特征向量都输入目标神经网络进行特征提取,得到对应的12个声纹特征向量;

[0093] A3: 依次取12个声纹特征向量中的一个声纹特征向量,计算其它声纹特征向量与当前的该声纹特征向量的相似度,若相似度最高的声纹特征向量与该声纹特征向量同属一个用户,则判断模型识别正确,否则识别错误;重复执行直至遍历完12个声纹特征向量;

[0094] A4: 统计步骤A3的识别结果,得到最终的准确率。

[0095] 经验证,本申请实施例的声纹识别方法的准确率比通过GE2E进行声纹识别的方法的准确率高。示例性地,在一次测试结果中,采用GE2E网络的声纹识别方法的准确率为0.704,采用本申请实施例的目标神经网络进行声纹识别的准确率为0.805。

[0096] (二) 运算速度测试

[0097] B1: 从数据集中获取来自6个用户、每人2句语音的样本数据,即 $6 \times 2 = 12$ 个样本数据(可以为12个numpy文件)作为一个批次输入到目标神经网络中进行测试,并记录目标神经网络运行消耗的时间;

[0098] B2: 重复B1步骤100次,得到100个耗时数据,去除其中的最大值和最小值,得到剩下的98个耗时数据,并对这98个耗时数据求均值作为最终耗时,并与基于GE2E的声纹识别方法的运行耗时进行比对。

[0099] 经过比对,本申请实施例通过该目标神经网络进行声纹识别的方法的最终耗时低于基于GE2E的声纹识别方法的运行耗时。示例性地,在一次测试结果中,采用GE2E网络的声纹识别方法的运行耗时为0.656秒,采用本申请实施例的目标神经网络进行声纹识别的的运行耗时为0.040秒。

[0100] 实施例三:

[0101] 图4示出了本申请实施例提供的一种声纹识别装置的结构示意图,为了便于说明,仅示出了与本申请实施例相关的部分:

[0102] 该声纹识别装置包括:音频特征向量获取单元41、目标神经网络单元42、确定单元43。其中:

[0103] 音频特征向量获取单元41,用于获取待识别音频的音频特征向量,所述音频特征向量包括时间维度和频谱特征维度,所述时间维度上的一个单位时间对应所述频谱特征维度上的一组频谱特征信息。

[0104] 可选地,所述音频特征向量获取单元41包括待识别音频获取模块和MFCC特征提取模块:

[0105] 待识别音频获取模块,用于获取待识别音频,滤除所述待识别音频的静音,得到有效音频段;并按照目标时长截取所述有效音频段,得到目标音频;

[0106] MFCC特征提取模块,用于提取所述目标音频的梅尔倒谱系数MFCC特征,得到音频特征向量。

[0107] 目标神经网络单元42,用于将所述音频特征向量输入目标神经网络,得到所述音频特征向量对应的目标声纹特征向量,其中所述目标神经网络由SENet模块、空洞卷积网络和全连接层组成,所述空洞卷积网络包含若干层用于提取所述音频特征向量在时间维度上的上下文信息的空洞卷积层。

[0108] 可选地,所述目标神经网络单元包括训练模块,用于获取样本数据,其中所述样本数据来自不同用户的音频数据;将所述样本数据输入所述目标神经网络进行训练,直至类内音频相似度和类间音频相似度满足预设条件,得到训练后的目标神经网络;其中,所述类内音频相似度为属于同一用户的声纹特征向量之间的相似度,所述类间音频相似度为属于不同用户的声纹特征向量之间的相似度。

[0109] 可选地,所述训练模块,具体用于依次将预设样本数据输入所述目标神经网络进行训练,直至目标函数的值满足预设条件,得到训练后的目标神经网络,所述目标神经网络的目标函数为:

$$[0110] \quad S_c = \sum_i^{NM} \left(\sum_j^{M-1} \text{sim}\{(v_i, v_j) | i \neq j, v_i \in P, v_j \in P\} - \max(\text{sim}\{(v_i, v_k) | v_i \in P, v_k \notin P\}) \right)$$

[0111] 其中, S_c 为目标函数的值,表示类内音频相似度与类间音频相似度之间的差值; N 为当前批次输入的样本数据对应的用户的数目, M 为每个用户对应的样本数据数量, v_i 表示当前批次的任意一个样本数据经过目标神经网络模型得到的声纹特征向量, P 表示 v_i 对应的用户, v_j 为与 v_i 同属于一个用户的声纹特征向量, v_k 为与 v_i 不属于同一个用户的声纹特征向量, $\text{sim}\{(v_i, v_j) | i \neq j, v_i \in P, v_j \in P\}$ 表示 v_i 与同属于 P 的声纹特征向量 v_j 的余弦相似度, $\text{sim}\{(v_i, v_k) | v_i \in P, v_k \notin P\}$ 表示 v_i 与其他用户的声纹特征向量 v_k 的余弦相似度。

[0112] 可选地,所述训练模块包括学习率调整模块,用于根据预设的目标学习率和当前训练步数动态调整目标神经网络在训练时的学习率。

[0113] 可选地,所述声纹识别装置应用于远场录音场景,所述样本数据包括携带背景噪声的远场录音数据、预设数量的无噪声音频数据。

[0114] 可选地,所述目标神经网络具体由第一卷积层、SENet模块、第一重构层、第一全连

接层、第二重构层、空洞卷积网络、第三重构层、平均池化层和第二全连接层组成,所述目标神经网络单元42具体用于:

[0115] 将所述音频特征向量输入目标神经网络,通过所述第一卷积层,得到第一特征向量,其中所述第一特征向量包含时间维度、频谱特征维度和通道维度;

[0116] 所述第一特征向量通过所述SENet模块,对每个通道的信息进行加权,得到第二特征向量;

[0117] 所述第二特征向量依次通过所述第一重构层、所述第一全连接层及所述第二重构层,得到第三特征向量;

[0118] 所述第三特征向量通过所述空洞卷积网络,依次经过若干层空洞卷积层,提取所述第三特征向量在时间维度的上下文信息,得到第四特征向量,其中每层空洞卷积层包含一个尺寸为 $n \times 1$ 的卷积核, n 为大于1的正整数,“ \times ”为乘号;

[0119] 所述第四特征向量依次通过第三重构层、平均池化层和第二全连接层,得到目标大小的目标声纹特征向量。

[0120] 确定单元43,用于将所述目标声纹特征向量与已注册的声纹特征向量相比较,确定所述待识别音频对应的目标用户。

[0121] 可选地,所述确定单元43还包括:

[0122] 指示模块,用于若未找到与所述目标声纹特征向量相匹配的已注册的声纹特征向量,则指示用户将所述目标声纹特征向量进行注册。

[0123] 需要说明的是,上述装置/单元之间的信息交互、执行过程等内容,由于与本申请方法实施例基于同一构思,其具体功能及带来的技术效果,具体可参见方法实施例部分,此处不再赘述。

[0124] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将所述装置的内部结构划分成不同的功能单元或模块,以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中,上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。另外,各功能单元、模块的具体名称也只是为了便于相互区分,并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0125] 实施例四:

[0126] 图5是本申请一实施例提供的终端设备的示意图。如图5所示,该实施例的终端设备5包括:处理器50、存储器51以及存储在所述存储器51中并可在所述处理器50上运行的计算机程序52,例如声纹识别程序。所述处理器50执行所述计算机程序52时实现上述各个声纹识别方法实施例中的步骤,例如图1所示的步骤S101至S103。或者,所述处理器50执行所述计算机程序52时实现上述各装置实施例中各模块/单元的功能,例如图4所示单元41至43的功能。

[0127] 示例性的,所述计算机程序52可以被分割成一个或多个模块/单元,所述一个或者多个模块/单元被存储在所述存储器51中,并由所述处理器50执行,以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机程序指令段,该指令段用于

描述所述计算机程序52在所述终端设备5中的执行过程。例如,所述计算机程序52可以被分割成音频特征向量获取单元、目标神经网络单元及确定单元,各单元具体功能如下:

[0128] 音频特征向量获取单元,用于获取待识别音频的音频特征向量,所述音频特征向量包括时间维度和频谱特征维度,所述时间维度上的一个单位时间对应所述频谱特征维度上的一组频谱特征信息。

[0129] 目标神经网络单元,用于将所述音频特征向量输入目标神经网络,得到所述音频特征向量对应的目标声纹特征向量,其中所述目标神经网络由SENet模块、空洞卷积网络和全连接层组成,所述空洞卷积网络包含若干层用于提取所述音频特征向量在时间维度上的上下文信息的空洞卷积层。

[0130] 确定单元,用于将所述目标声纹特征向量与已注册的声纹特征向量相比较,确定所述待识别音频对应的目标用户。

[0131] 所述终端设备5可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述终端设备可包括,但不限于,处理器50、存储器51。本领域技术人员可以理解,图5仅仅是终端设备5的示例,并不构成对终端设备5的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件,例如所述终端设备还可以包括输入输出设备、网络接入设备、总线等。

[0132] 所称处理器50可以是中央处理单元(Central Processing Unit,CPU),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0133] 所述存储器51可以是所述终端设备5的内部存储单元,例如终端设备5的硬盘或内存。所述存储器51也可以是所述终端设备5的外部存储设备,例如所述终端设备5上配备的插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)等。进一步地,所述存储器51还可以既包括所述终端设备5的内部存储单元也包括外部存储设备。所述存储器51用于存储所述计算机程序以及所述终端设备所需的其他程序和数据。所述存储器51还可以用于暂时地存储已经输出或者将要输出的数据。

[0134] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将所述装置的内部结构划分成不同的功能单元或模块,以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中,上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。另外,各功能单元、模块的具体名称也只是为了便于相互区分,并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0135] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述或记载的部分,可以参见其它实施例的相关描述。

[0136] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单

元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0137] 在本申请所提供的实施例中,应该理解到,所揭露的装置/终端设备和方法,可以通过其它的方式实现。例如,以上所描述的装置/终端设备实施例仅仅是示意性的,例如,所述模块或单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以通过一些接口,装置或单元的间接耦合或通讯连接,可以是电性,机械或其它的形式。

[0138] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0139] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0140] 所述集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请实现上述实施例方法中的全部或部分流程,也可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是,所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减,例如在某些司法管辖区,根据立法和专利实践,计算机可读介质不包括电载波信号和电信信号。

[0141] 以上所述实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围,均应包含在本申请的保护范围之内。

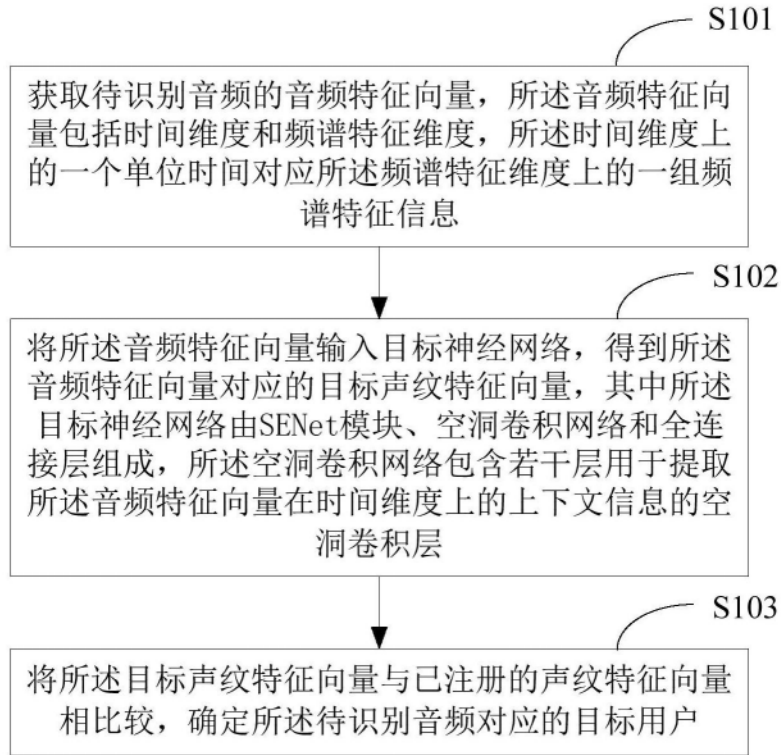


图1

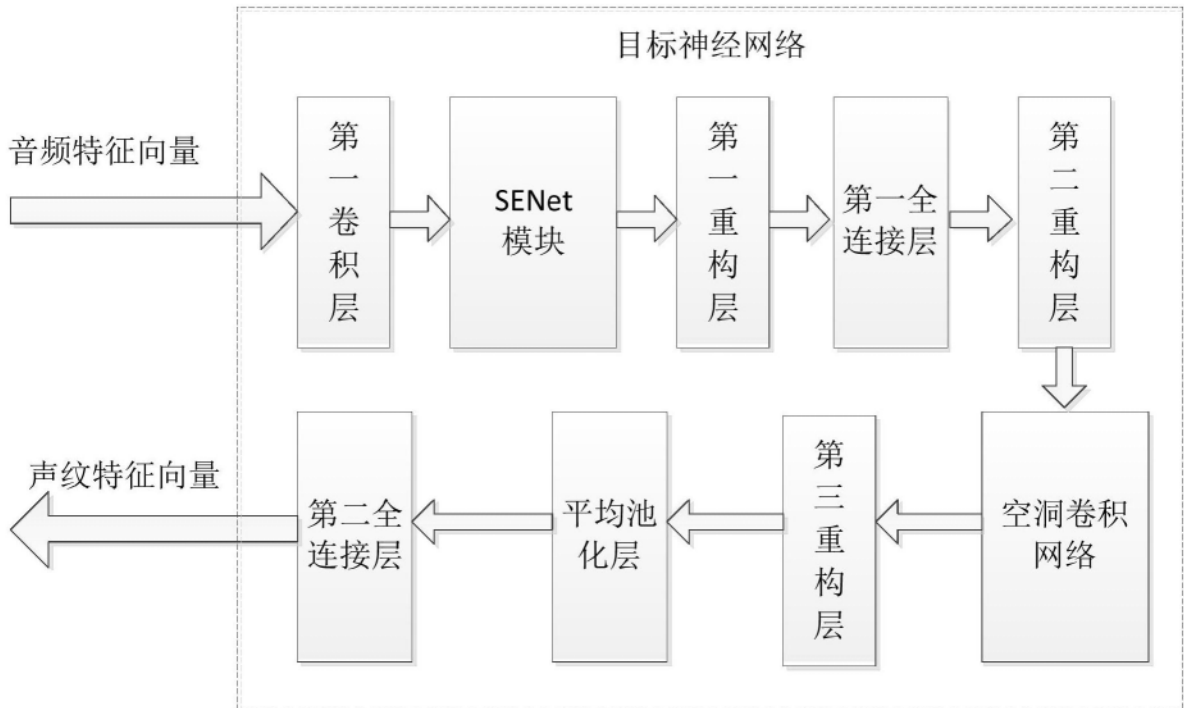


图2

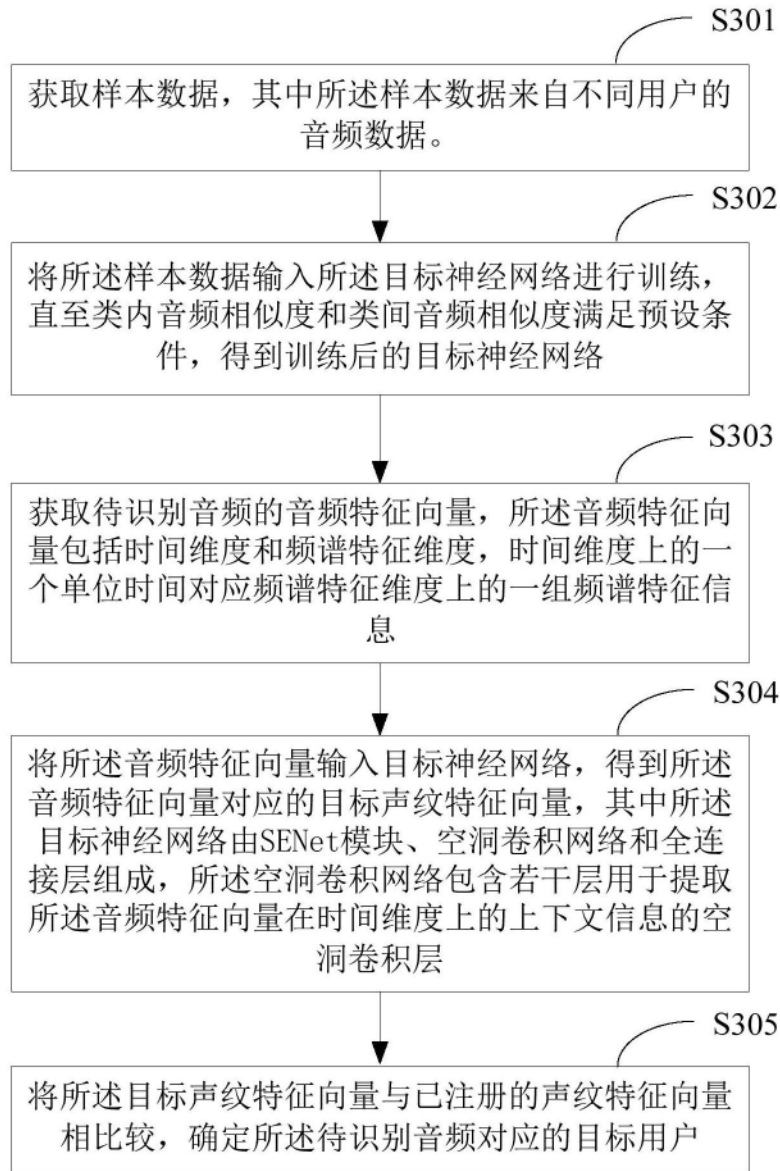


图3

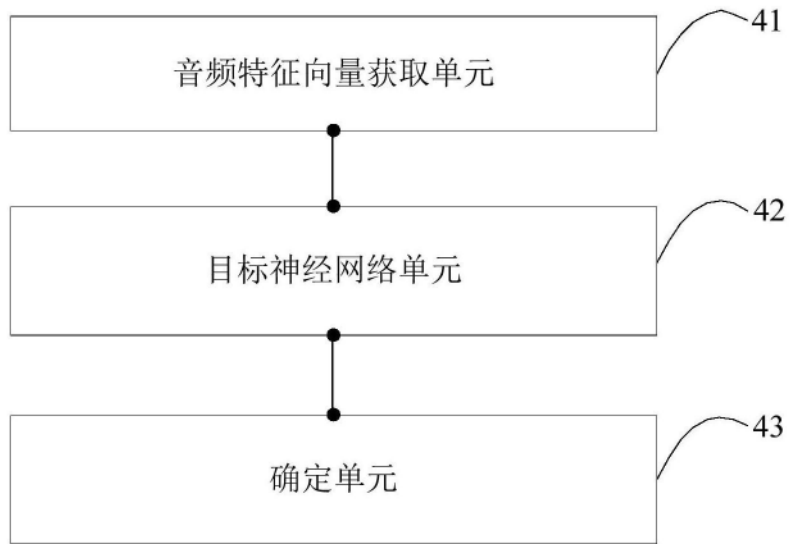


图4

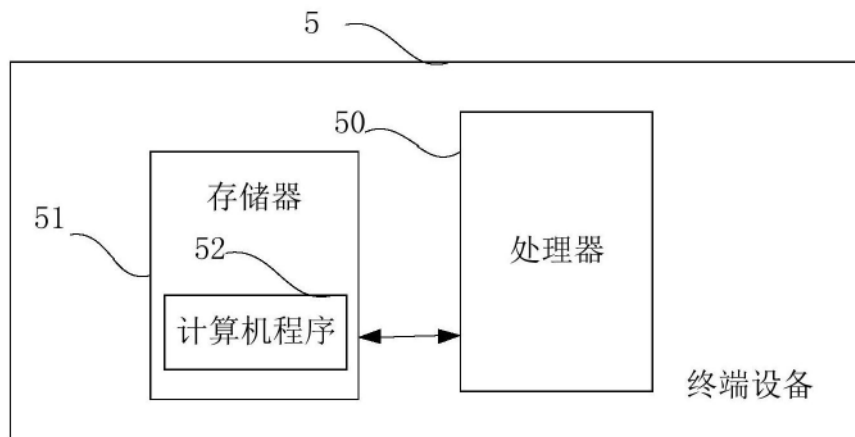


图5